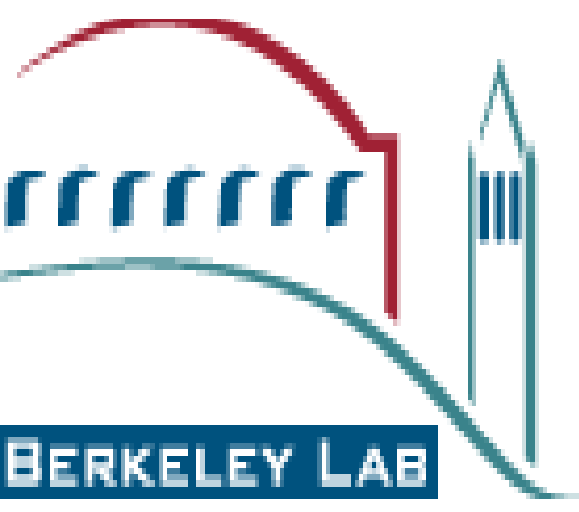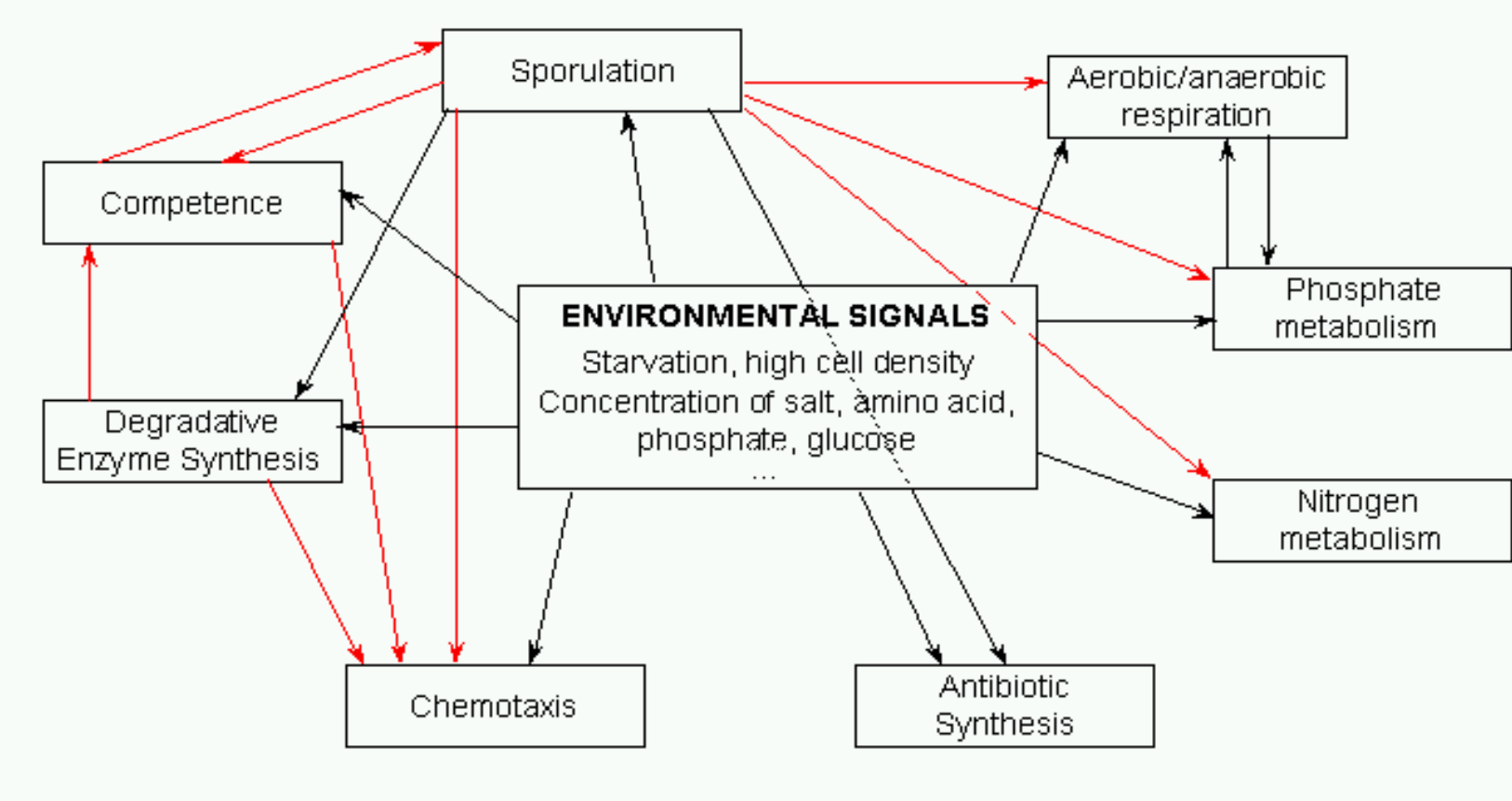# Pathway Evolution and Modularity of Bacterial Stress Responses

Amoolya Singh[1], Denise M. Wolf[2], and Adam P. Arkin[2,3,4]

[1]Computer Science Division, University of California – Berkeley, [2]Physical Biosciences Division, Lawrence Berkeley National Lab,
[3]Bioengineering Department, University of California – Berkeley, [4]Howard Hughes Medical Institute

Motivating Questions

How do biological networks evolve over time?

- *Network*: loosely defined as biochemical, signal transduction, regulatory, or developmental network of genes, proteins, and metabolites.
- *Stress response network*: heat shock, osmotic shock, metal ion uptake, oxidative stress / anaerobiosis, chemotaxis, competence development, antibiotic synthesis.

Are certain genes in a developmental pathway more mutable than others? which ones? why?

How much of network arose independently, how much was laterally transferred?

Did pathways arise together? separately? in what order?

Does variability in pathway gene content between various bacteria correlate with environmental niche? with function?

Any network structure evident in these patterns of evolution? identifiable modules conserved and/or laterally transferred? what does each module contain?

Methods

Comparative genomics: To identify gene content, order & homology.

List all species exhibiting stress responses of interest

Focus on species with fully sequenced genomes

List of all genes implicated in each stress response

Construct ortholog set for each gene using 3-way bidirectional best hits; examine size and diameter of ortholog set

Phylogenetics: To trace vertical evolution of a group of genes.

Construct a phylogenetic tree for each protein

Do set of protein trees agree with well-known 16S rRNA tree? Did they evolve at the same rate? Were they inherited together as a group?

Genomic & Phylogenetic Analysis: To trace lateral gene transfer; look for

unexpected ranking in sequence similarity or tree topology

anomalous nucleotide composition (GC content, codon usage bias)

insertion/deletion sites; evidence of viral DNA

Statistical Analysis: To observe patterns of conservation of a group of genes using hierarchical clustering of genes and species; cluster by:

phylogenetic profile, known species inter-relations, ecological niche, phenotype, morphology, type of primary metabolism

Use singular value decomposition and other multivariate statistical techniques to filter noise and infer relationships between evolution and function
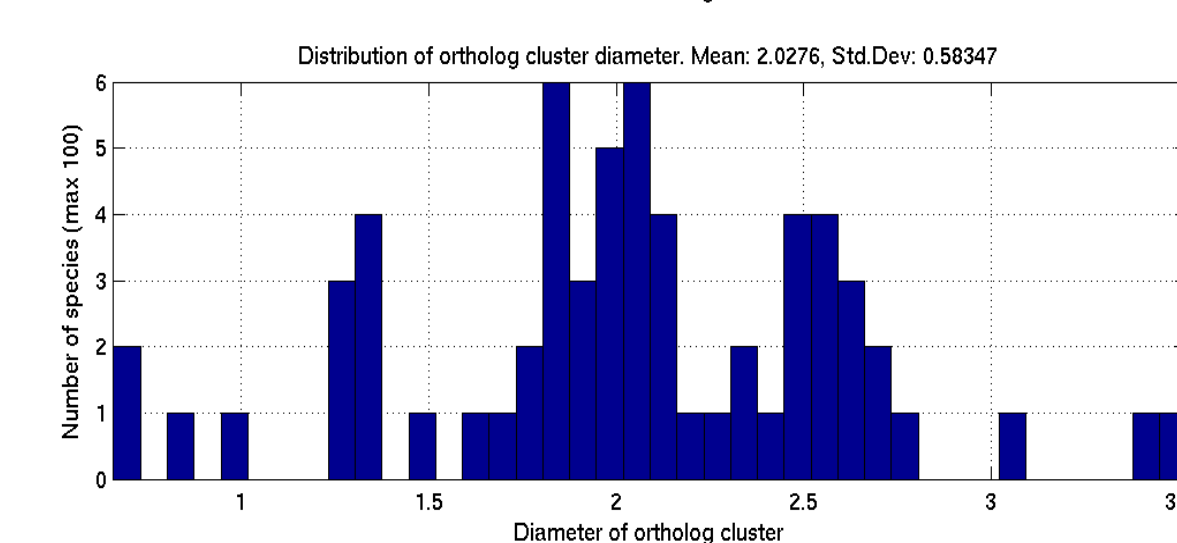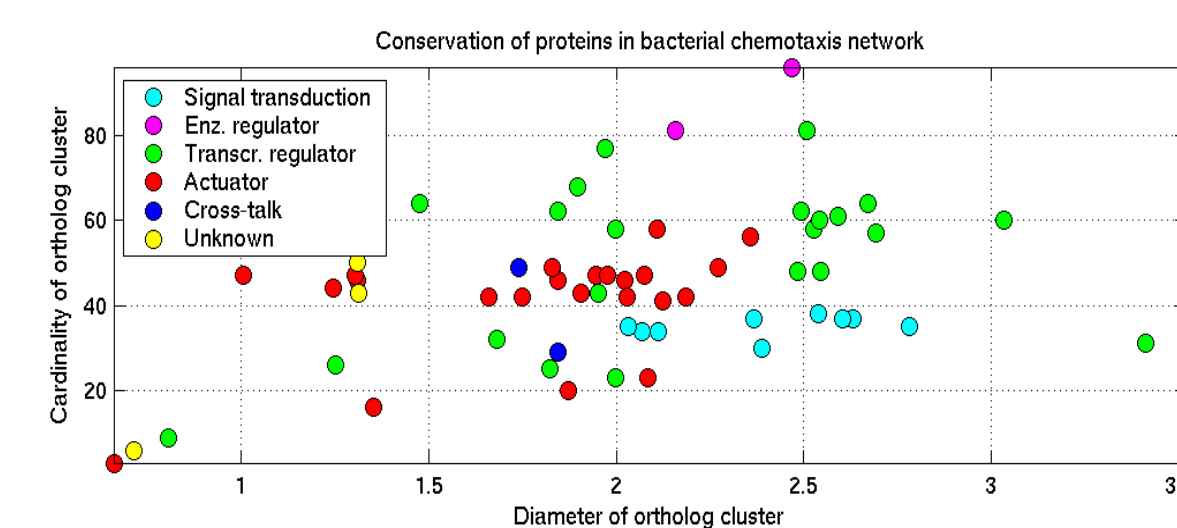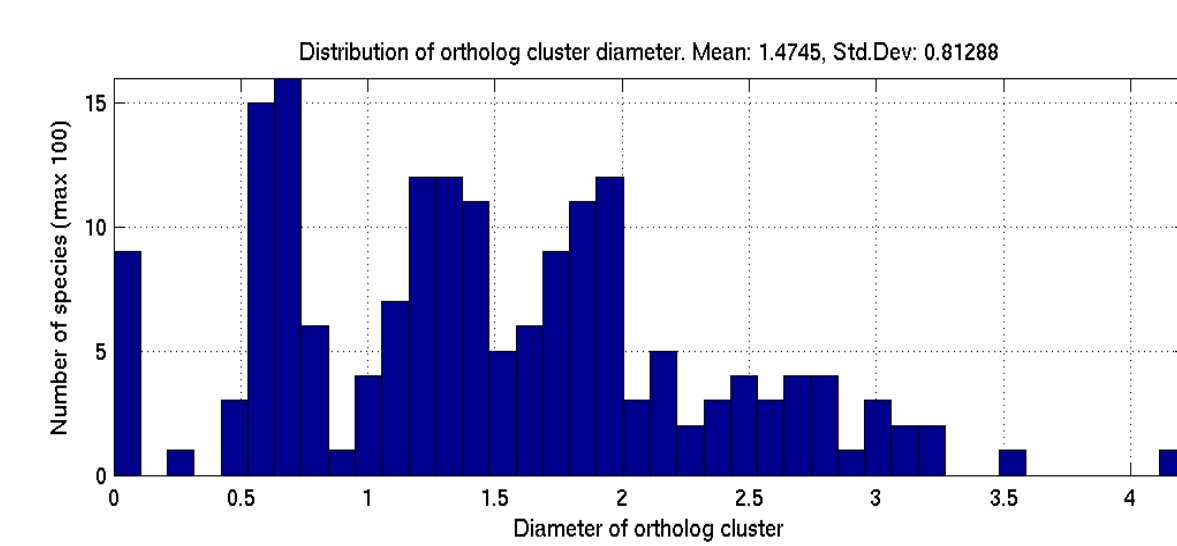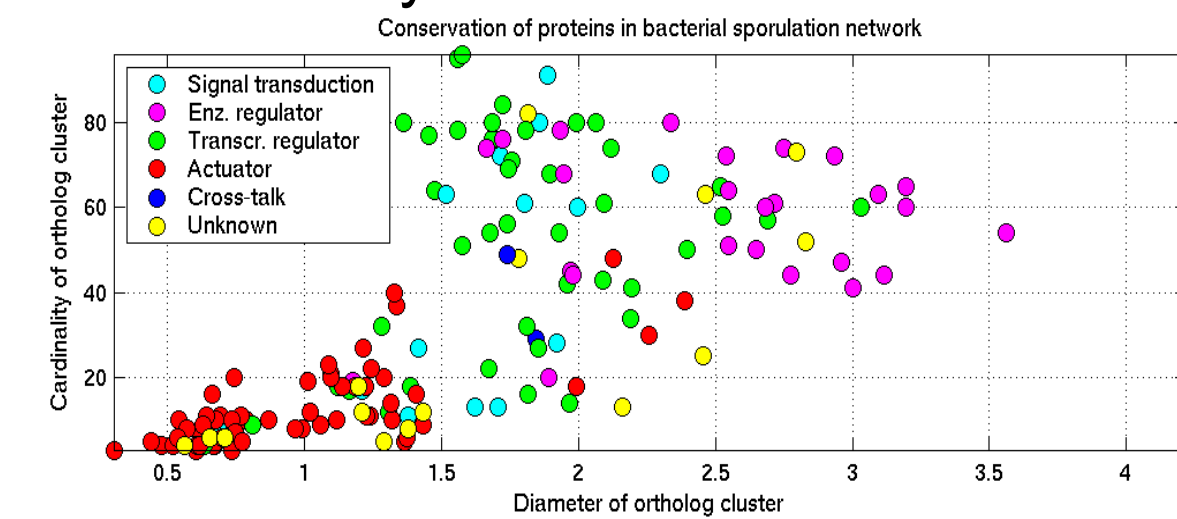
Results

Phylogenetic trees of some spore-forming bacteria.

Consensus of trees constructed from various proteins in the sporulation pathway does not agree with well-known tree constructed from rRNA. Although they participate in the same biochemical pathway, these proteins probably evolved at differing rates and have disparate origins.
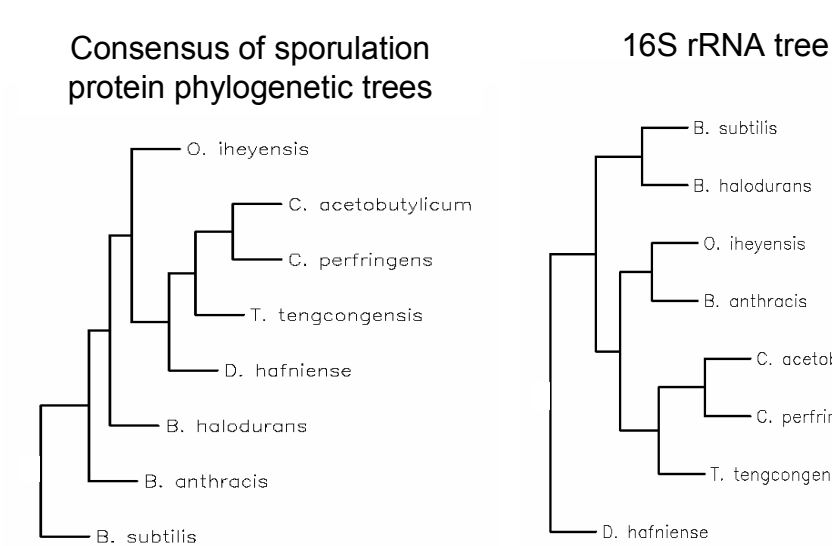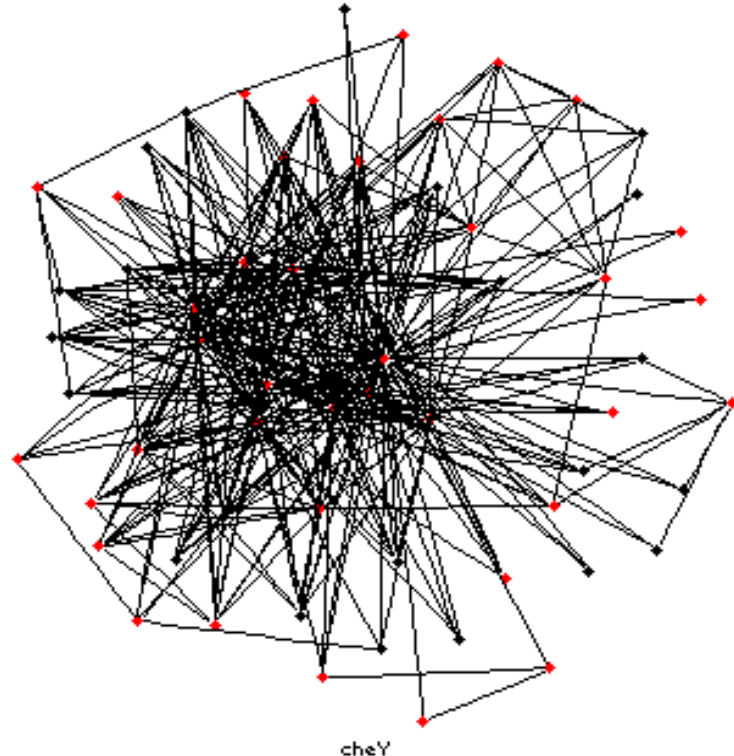

Different parts or "modules" of a gene regulatory network have different evolutionary histories.

In both sporulation and chemotaxis networks, functional modules correspond to evolutionary modules.

Example ortholog cluster for CheY, major response regulator protein in chemotaxis pathway.
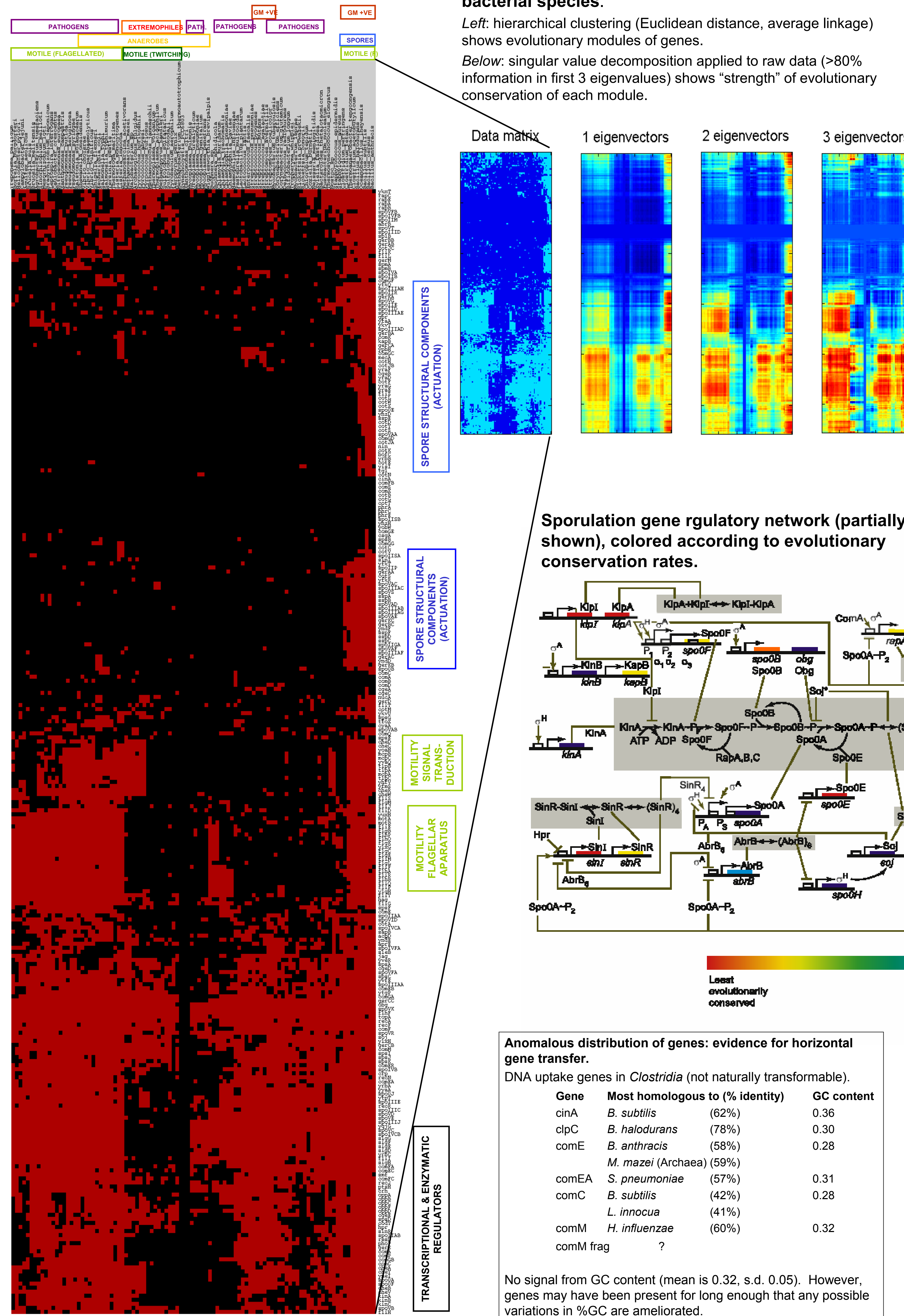62 nodes (species), diameter 2.49, motile species in red.

Distribution of 278 stress response genes in 100 bacterial species.

Left: hierarchical clustering (Euclidean distance, average linkage) shows evolutionary modules of genes.

Below: singular value decomposition applied to raw data (>80% information in first 3 eigenvalues) shows "strength" of evolutionary conservation of each module.

Sporulation gene rgulatory network (partially shown), colored according to evolutionary conservation rates.

Anomalous distribution of genes: evidence for horizontal gene transfer.

DNA uptake genes in *Clostridia* (not naturally transformable).

Gene	Most homologous to	(% identity)	GC content
cinA	*B. subtilis*	(62%)	0.36
clpC	*B. halodurans*	(78%)	0.30
comE	*B. anthracis*	(58%)	0.28
	M. mazei (Archaea)	(59%)	
comEA	*S. pneumoniae*	(57%)	0.31
comC	*B. subtilis*	(42%)	0.28
	L. innocua	(41%)	
comM	*H. influenzae*	(60%)	0.32
comM frag	?		

No signal from GC content (mean is 0.32, s.d. 0.05). However, genes may have been present for long enough that any possible variations in %GC are ameliorated.

Discussion

Chemotaxis network is modular. Presence of flagellar apparatus and associated regulators distinguishes motile from non-motile bacteria.

Sub-modules: archaeal & type IV twitching motility; other eubacterial systems of motility. Presence of specific global regulators distinguishes these modules.

Sporulation less modular; genes in pathway appear to have evolved at different rates and co-opted from different sources. Possibly due to close entanglement with the DNA replication and cell division apparatus.

Conservation of sporulation genes correlates clearly with function. *Most highly & widely conserved*: global regulators and sigma factors. *Highly conserved*: stages III-V sporulation genes (the core process of making a spore). *Not conserved*: small enzymes, coat proteins, and pheromone systems.

Competence evolved multiple times. At least three distinct systems of DNA uptake appear to exist: gram-positive, gram-negative, and the unique system in *Streptococcus*.

Future Work

Comparative dynamics

Network structure

Graph theoretic analysis

Conservation of regulatory sites

References

A. Singh, D. M. Wolf, and A.P. Arkin. (2003) Pathway Evolution and Modularity of Bacterial Stress Responses. 103rd General Meeting, American Society for Microbiology.

D. M. Wolf and A. P. Arkin. (2003) Motifs, Modules, and Games in Bacteria. *Current Opinion in Microbiology* 6(2) 125 –134.